

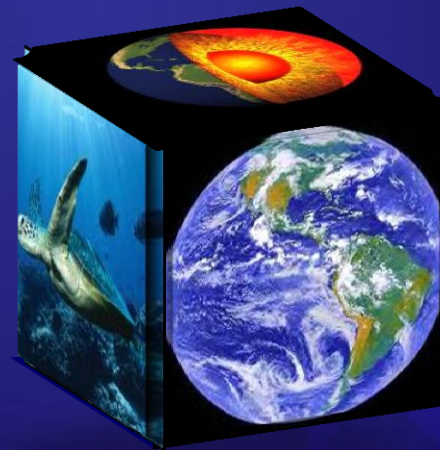
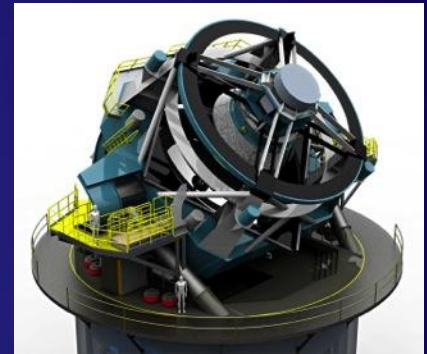
Scientific Data: Changing and Changing The World

Rob Pennington

Program Director

Office of Cyberinfrastructure

NSF



The Shift Towards a "Sea of Data"

Implications



❖ All science is moving towards a data-dominated world

- Fundamental questions become focused around data: How to remove boundaries? How to incentivize sharing?

❖ Software

➤ Publications

How do we attribute credit for this new publication form? How are data peer reviewed? What is a publication in the modern data-rich world?

❖ Totally new methodologies

- Algorithms, mathematics, culture

❖ Data become the medium for

- Multidisciplinarity, communication, publication...science



Recent Data Related Activities

- ❖ National Science Board Task Force on Data Policies Recommendations
 - http://www.nsf.gov/nsb/committees/tskforce_dp.jsp
- ❖ US Office of Science & Technology Policy
 - <http://www.whitehouse.gov/blog/2011/11/07/request-information-public-access-digital-data-and-scientific-publications>
 - “Request for Information on Public Access to Digital Data and Scientific Publications”
 - “Public Access to Digital Data Resulting From Federally Funded Scientific Research”
- ❖ Coordinated US-EU programs
 - INFRA 2012-3.1, 3.2 and OCI/GEO/MPS DCL (STCI)⁶

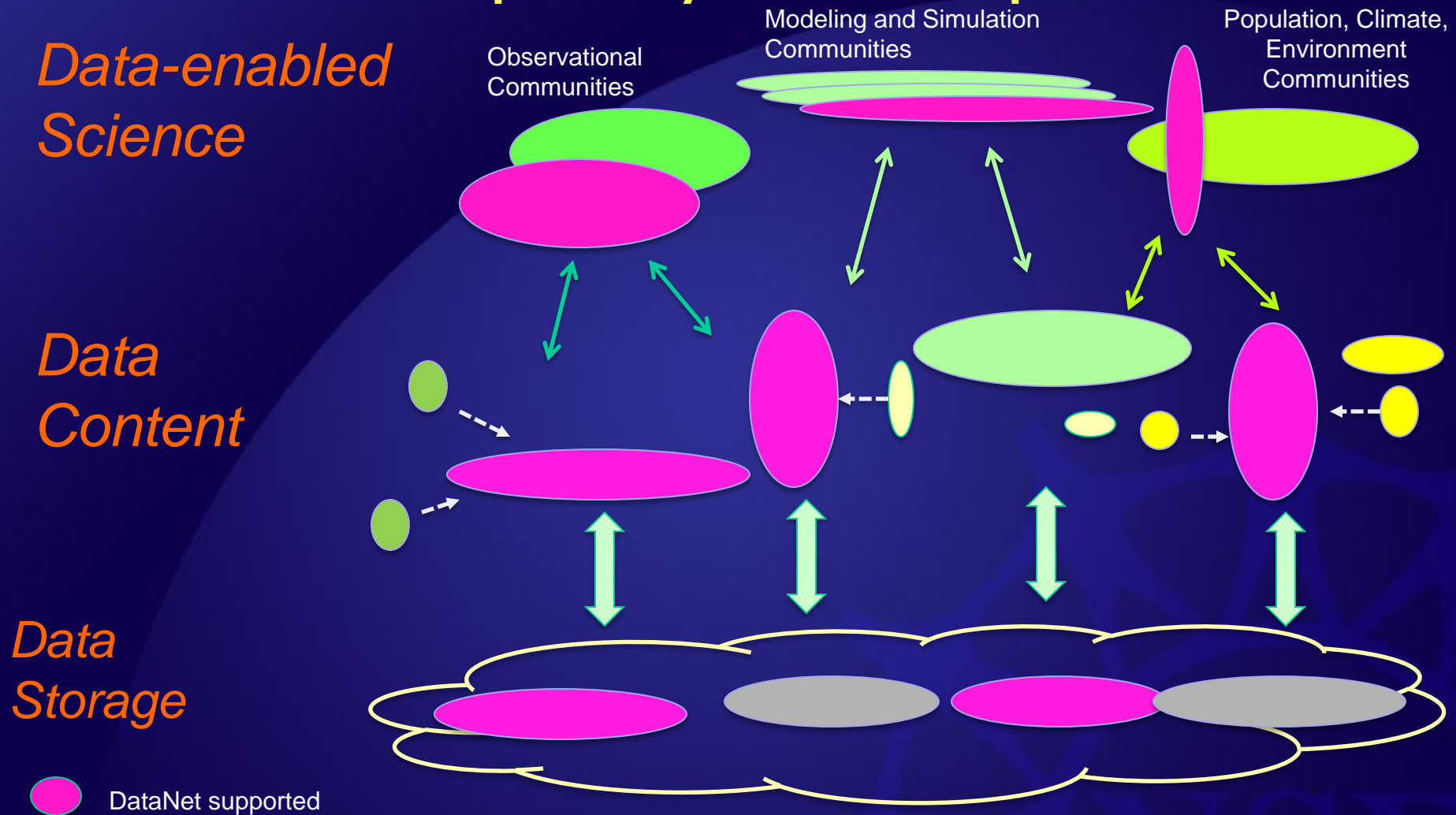
Changes Coming at NSF for Data!

- ❖ Data are becoming:
 - Primary means of communication through sharing
 - Major product of research (including publication)
- ❖ Long-standing NSF Policy on Data:
 - *"Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data... created or gathered in the course of work under NSF grants"*
- ❖ NSF now requires a Data Management Plan (DMP):
 - 2-page supplement to the proposal
 - Subject to peer review; criterion for award
 - Not possible to submit proposals without DMP

Data-Enabled Science

- ❖ Data Services Program (*data*)
 - Provide reliable digital preservation, access, integration, and analysis capabilities for science and/or engineering data over a decades-long timeline
- ❖ Data Analysis and Tools Program (*information*)
 - Data mining, manipulation, modeling, visualization, decision-making systems
- ❖ Data-intensive Science Program (*knowledge*)
 - Intensive disciplinary efforts, multi-disciplinary discovery and innovation

DataNet: A Multi-tiered and Multi-Disciplinary Landscape



Summary

- ❖ Think about effective ways to approach the challenges associated with data
 - Critical concepts and goals
 - Realistic and innovative
- ❖ Structure for longevity
 - Scalable open inclusive governance
 - Long term business models